Transposons continue the amaze

Nermin Gozukirmizi

1Department of Molecular Biology and Genetics, Faculty of Art and Science, Istinye University, Istanbul/Turkey

Abstract

Transposable elements (TEs) were first discovered in maize plants. However, they exist almost in all species with a few exceptions (Plasmodium falciparum, Ashbya gossypii and Kluyveromyces lactis). They are the most important contributors to genome plasticity and evolution and even epigenetic genome regulation. Organisms with large genomes have high transposon percentages. For example, Arabidopsis thaliana has a genome size of 125 Mb which comprises 14% transposons, Homo sapiens (3000 Mb) 45-48.5%, and Hordeum vulgare genome (5300 Mb) has 80%. TEs are classified into two major groups based on their transposition mechanisms: Class I (RNA transposons – retrotransposons) and Class II (DNA transposons). Recent progress in whole-genome sequencing and long-read assembly have resulted in identification of unprecedentedly long transposable units spanning dozens or even hundreds of kilobases, initially in prokaryotic and more recently in eukaryotic systems. All TEs in a cell are named as transposome (mobilome), and transposomics is a new area to work with transposome. Although a number of bioinformatics softwares have recently been developed for the annotation of TEs in sequenced genomes, there are very few computational tools strictly dedicated to the identification of active TEs using genome-wide approaches. In this review article, after a brief introduction and review of the transposable elements, I discussed their effects in gene expression, evolution, recent applications and also share our research on retrotransposons with different organisms.

1. Introduction

Transposon, a segment of DNA that moves to a new location in a chromosome, to another chromosome or cell, even different organism and alters existing genetic structure, causes significant changes and play an important role during evolution. Transposons were first described by Barbara McClintock (1950), a maize cytogeneticist who was rewarded with
Nobel prize almost 30 years later than her exploration of the relationship between chromosome breaks and maize grain colour alterations. Today, we are aware that gene and genome plasticity caused by transposons exist in somatic tissues almost all living organisms. Different terms, such as jumping genes, mobile elements, controlling elements, transposable elements are used in synonymous ways. Although there are a vast amount of knowledge on structure, types and life cycles of these genomic sequences everyday with the recent progress in whole-genome sequencing and long-read assembly, combined with expansion of the familiar range of model organisms, resulted in new informations for their origin, functions, roles in gene expression, and evolutionary processes. TEs in the genomes of living organisms, are either defective, fossilized copies or potentially active copies that are restrained by host silencing systems. However, active transposition evidenced by instances of mutagenic (yet potentially evolutionarily significant) insertions has been demonstrated. For example, TEs have been shown to silence or alter expression of genes adjacent to insertion sites, contribute to chromosomal rearrangements via recombination, epigenetically alter regional methylation patterns, and provide template sequences for RNA interference (Feschotte et al., 2002; Bennetzen, 2005; Morgante et al., 2007; Weil and Martienssen, 2008; Slotkin et al., 2012; Lerat et al., 2019). This diverse functional impacts of TEs, and their intrinsic contribution to genomic plasticity suggest that these elements play a major role in molecular diversification, and ultimately, species divergence (Schrader and Schmitz, 2019; Dupeyron et al., 2019; Boissinot et al., 2019). This review article covers a short overview of TEs classification, transposition mechanisms, their effects on gene expression and evolutionary processes and even importance and usage of transposons for different purposes such as transposon markers, transposomics, and . Particularly their impact on protein coding and those TE-derived small RNAs have potential to affect non-TE transcripts by sequence complementarity, thereby generating novel gene regulatory networks including stress resistance and hybridization barrier. Apart from the small RNAs, number of long non-coding RNAs (lncRNAs) are originated from TEs in plants (Cho et al., 2018).

2. Structure of Transposons

According to the traditional simplistic representation transposons are classified into two class as Class I and Class II (retrotransposons and DNA transposons, respectively). However, there are many subtypes under this classification (Piégua et al., 2015) (Figure 1). Nowadays over-sized transposons come into consideration with some of the proposed models for gene
capture by eukaryotic TEs (Arkhipova et al., 2019). Before we mention about types and classification of transposons, general explanation of their structures will be helpful to understand the following parts of this review. Transposons use many different enzymes for their transposition. While some transposons can encode these enzymes (autonomous), others can not (non-autonomous) and use enzymes of autonomous transposons.

Figure 1 indicated comparison and content from two proposals for the classification and annotation of eukaryotic TEs. The Repbase proposal is shown on the right (Jurka et al., 2005; Kapitonov and Jurka, 2008) and the Wicker proposal on the left (Wicker et al., 2007). Both proposals are based on DNA and amino acid sequence features. Both proposals divide all TEs into two groups, the retrotransposons and the DNA transposons (Piégua et al., 2015).

<table>
<thead>
<tr>
<th>Wicker’s proposition</th>
<th>Repbase proposition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classification</strong></td>
<td><strong>Classification</strong></td>
</tr>
<tr>
<td>Order</td>
<td>Superfamily</td>
</tr>
<tr>
<td>LTR</td>
<td>Copia</td>
</tr>
<tr>
<td>Gypsy</td>
<td></td>
</tr>
<tr>
<td>Bal-Pao</td>
<td></td>
</tr>
<tr>
<td>Retrotransposons</td>
<td></td>
</tr>
<tr>
<td>DIRS</td>
<td>DIRS</td>
</tr>
<tr>
<td>Ngao</td>
<td></td>
</tr>
<tr>
<td>PLE</td>
<td>PLE</td>
</tr>
<tr>
<td>LINE</td>
<td>R2</td>
</tr>
<tr>
<td>RTE</td>
<td>RT</td>
</tr>
<tr>
<td>Jockey</td>
<td>Jockey</td>
</tr>
<tr>
<td>L1</td>
<td></td>
</tr>
<tr>
<td>SINE</td>
<td>SINE</td>
</tr>
<tr>
<td>TE1</td>
<td>TE1</td>
</tr>
<tr>
<td>S5</td>
<td>S5</td>
</tr>
</tbody>
</table>

| Figure 1. New classification of transposons. |
Retrotransposons have more complex structure than DNA transposons. They have LTR (Long Terminal Repeat), R (Repeated region), U3 (Unique for 3’end of RNA), U5 (Unique for 5’ end of RNA), PBS (Primer Binding Site), GAG (Group-specific antigen), POL (Polyprotein), AP (Aspartic Peptidase), RT (Reverse Transcriptase), RH (Ribonuclease H), INT (Integrase), ENV (Envelope), PPT (PolyPurine Tract) and TSD (Target Site Duplication) (Figure 2).

![Figure 2. Schematic demonstration of a retrotransposon having LTR regions.](image)

Despite the complex structure of retrotransposons, DNA transposons have more simple structure (Figure 3). DNA transposons encode a transposase (Tase) enzyme. This enzyme cuts DNA transposon and integrates it to a new location.

![Figure 3. Schematic demonstration of a DNA transposon.](image)

(TIR, Terminal Inverted Repeat; Tase, Transposase; TSD, Target Site Duplication).

DNA transposons also contain non-coding repeat regions, like LTR, at both borders. But these repeats are present as inverted orientation (Poulter and Goodwin, 2005). Thus, these repeats are called Terminal Inverted Repeats (TIRs). TIRs are essential for the transposition of most transposons. Generally, it is a characteristic feature of DNA transposons. However, some retrotransposon groups also have these repeat sequences. Generally two functional regions are present in TIRs sequences. While the first region is involved in DNA cleavage and strand transfer reactions, second one is required for specific recognition and binding (Szabo et al., 2010) (Figure 3). Over-sized transposons; recently described by Arkhipova and Yushenova (2019). They produced in the genome generally end bypass, hybrid end formation, secondary nested insertion and chimeric retrotranscript formation (Figure 4).
3. Transposon Markers

A molecular marker is defined as a particular segment of DNA that is representative of the differences at the genome level. An ideal molecular marker should be polymorphic and evenly distributed throughout the genome, generate multiple, independent and reliable markers, simple, quick, and inexpensive, need small amounts of DNA samples etc (Agarwal et al., 2008). Many features of LTR retrotransposons in plant genomes have made them excellent sources of molecular markers (Kalendar and Schulman, 2006; Poczai et al., 2013). Schematic demonstration of some LTR-retrotransposon based marker techniques are given at Figure 5. Retrotransposon sequences alone or combined with various sequences in the genome provide primer binding sites. LTRs of a retrotransposon have conserved sequences between different organisms. Thus a primer pair designed for a specific organism can be used for other organisms. However, different retrotransposons of an organism have different LTR sequences so a primer pair designed for a specific retrotransposon, can not bind to another’s

Figure 4. Production of over-sized transposons (Arkhipova et al., 2019).

1. End bypass (3’-transduction)

2. Hybrid end formation (5’ or 3’)

3. Secondary (nested) insertion (5’ or 3’)

4. Chimeric retrotrancripts (class I)
LTR. Inter-Retrotransposon Amplified Polymorphism (IRAP), Retrotransposon-Microsatellite Amplified Polymorphism (REMAP), Retrotransposon-Based Insertional Polymorphism (RBIP), Sequence-Specific Amplified Polymorphism (S-SAP), RAPD-Retrotransposon Amplified Polymorphism (R-RAP) are the main PCR based marker systems which are used successfully for genetic polymorphism, fingerprinting, developmental biology and evolutionary research.

Figure 5. LTR-retrotransposon based marker techniques. (a) IRAP (b) REMAP (c) RBIP (d) SSAP (e) RRAP (f) IPBS (Gozukirmizi et al., 2015).

4. Sequencing and Transposomomic

There are thousands of genomes are being sequenced currently. The Earth Biogenome Project (https://www.earthbiogenome.org) aim at sequencing large number of genomes. This expansion of genomic data creates an urgent need for modern software tools to aid in detecting LTR-RTs in the new sequenced genomes; such tools should remedy the limitations of the currently available tools. There are recently developed TE annotations systems such as RepetDB (Amselem et al., 2019) is designed to be a TE knowledge base populated with full de novo TE annotations of complete (or near-complete) genome sequences. Indeed, the description and classification of TEs facilitates the exploration of specific TE families, superfamilies or orders across a large range of species. It also makes possible cross-species searches and comparisons of TE family content between genomes. LtrDetector (Valencia and
5. Transposons and Gene Expression

Transposable Elements (TEs) have been shown to alter gene regulation and effects genome evolution (Kazazian, 2004; Buckley and Adelson, 2014; Kapusta et al., 2017; Zeng et al., 2018; Deniz et al., 2019). TEs can exert these effects on genes by altering chromatin structure, providing novel promoters or insulators, novel splice sites or other post-transcriptional modifications to re-wire transcriptional networks important in development and reproduction.

Advances in genomics and epigenomics have brought in a new era in the study of TE regulation and its impact on host genomes. This includes investigating the roles of an expanding repertoire of DNA modifications that are potentially far more widespread across species than previously thought. Using the regulatory signals provided by DNA modifications to control TEs seems to have been a commonly adopted strategy throughout evolution, albeit displaying intriguingly high variation across even closely related species. Comparative genomics and epigenomics efforts will continue to provide clues into the intricate relationships between TE evolution and that of DNA-modifying enzymes. It is now clear that the impact of TEs on genomes is dictated to a large extent by the regulatory activities that target them, including the action of DNA-modifying enzymes. Heterochromatin is commonly regarded as silent DNA. It consists of large regions of repetitive nucleotide sequences and transposons. Transposons, however, must be suppressed because they constitute two dangers for the genome: (1) their repeated units can cause spurious homologous recombination; and their ability to transpose can lead to disruption or misregulation of important genes. Both of these dangers are suppressed by heterochromatinization (Madlung and Comai, 2004). Recent report that described a class of bacterial Tn7-like transposons encoding evolutionarily linked CRISPR–Cas systems and proposed a functional relationship between RNA-guided DNA targeting and transposition (Klompe et al., 2019). These recent technologies allows us to change the genome chance the gene expression in a directed manner.
6. Transposons and Evolution

Transposons, especially retrotransposons, play an important role in the evolution of living organisms. Despite being at very high proportions in the genome (45% in human, 90% in barley), most of the transposons are inactive or immobile in the genome. However, they can be activated and increased their movement by biotic and abiotic stresses such as salinity, medicines, herbicides and tissue culture conditions (Yilmaz et al., 2018. Retrotransposons are also important in terms of evolution, that retrotransposons found in plants are also could be found in the human genome. TE content strongly correlates with genome size variation. TE-derived DNA mostly makes up the 20-30% of the genome even for plant species with small genomes like Brachypodium distachyon and Arabidopsis spp. (The Arabidopsis Genome Initiative, 2000; The International Brachypodium Initiative, 2010). Mutator transposable elements (TEs) are among the most mutagenic transposons known, due to their very high rates of transposition and their bias for inserting near or close to genes (Dupeyron et al., 2019). More than any other genome components, TEs have the capacity to move across species barriers through Horizontal Transfer (HT), with substantial evolutionary consequences (Reiss et al., 2019). Most eukaryotic organisms accommodate high numbers of retrotransposons in centromeres and telomeres. Plant centromeres contain infused TEs within short centromeric repeats (Ma et al., 2007). Pericentromeric regions, similarly, are composed mostly of silenced TEs and pseudogenes (Hall et al., 2006). On the other hand, the exact role of these sequences in centromere or telomere function is still not clear (Joly-Lopez and Bureau, 2014). TEs are no longer only recognized as the negligible fraction of genomes, but they are considered as potential contributors to evolutionary adaptation. Following liberation from the host's silencing mechanisms, they might cause significant spontaneous changes. With the ongoing advances in sequencing technology, high-quality, long-read-based genome assemblies now begin to provide the foundation for studying the role of TEs in genome evolution and adaptations in an unprecedented matter.

7. Our Research with Transposons

Our group used retrotransposon-based molecular markers mainly for analysis of somaclonal variation and fingerprinting (Gozukirmizi et al., 2016). The stability of aging barley calli and callus-regenerated shoots was investigated by IRAP using primer derived from BARE-1 (Evrensel et al., 2011; Yilmaz and Gozukirmizi, 2013) and Nikita (Bayram et
al., 2012) sequences. Callus culture conditions activate BARE-1 and Nikita element. In the paper by Marakli et al. (2012), similarity of mature embryo, leaf and root tissues grown from the same barley plant were investigated in terms of BARE-1 and BAGY2 movements. BAGY2 was found to be more stable than BARE-1. Not all callus induction conditions increase retrotransposon activity (Temel and Gozukirmizi, 2013). However, in addition to BAGY2 retrotransposon specific IRAP polymorphism, they also caused increase in copy numbers of internal domains of BAGY2 (Yilmaz et al., 2014). Transformation of tobacco plant with dehE gene which degrades the herbicide Dalapon, was shown to cause activation of Tto1 retrotransposon, one of the few active retrotransposons in tobacco (Kaya et al., 2013). Moreover, we investigated non-autonomous retrotransposon Sukkula in barley (Kartal et al., 2014). Our group utilized IRAP technique to assess genotoxicity of some drugs e.g. epirubicine (Hamat-Mecbur et al., 2014) and amiprophos-methyl (Temel and Gozukirmizi, 2014). We report the soybean specific SIRE1 Retrotransposons in Barley Hordeum vulgare L. genome (Cakmak et al., 2015). Analysis on Hopi/Osr27 and Houba/Tos5/Osr13 retrotransposons in rice (Yuzbasioglu et al., 2016a, Yüzbaşıoğlu et al., 2016b) and in Pinus nigra barley specific BAGY2, Nikita and Sukkula were analysed (Marakli et al., 2019). We also observed barley specific retrotransposons in endemic Colchium chalcedonicum (Karlık et al., 2019). Recently we were able to find barley specific retrotransposons in human genome (Cakmak et al., 2017. We also tested retrotransposon movements for short term mutagenicity test (Yilmaz et al., 2018). We are planning develop universal short - term test system using plant retrotransposons.

8. Conclusion

TE are the main source of genome plasticity with recombination, somatic mutation and epigenetic factors. In the last decade, it became obvious that the mobilome was a fully valid participant of the genomic regulatory networks because of the formation of regulatory regions, noncoding transcripts, and domestication of TE-encoded proteins. Identification of the contribution of TEs to the uniqueness of each genome will be key to unraveling the impact of genome architecture on organismal evolution. At the same time, the mobilome evolution has specific features and takes place partly independently of the rest of the genome. Transposons are the major factor affecting the sizes of plant genomes. Under stressful conditions, they can rearrange a genome. TEs play roles in relocating genes and generating new genes and new pseudo genes. They can contribute to centromere function. TEs can
regulate the expression of nearby genes via several mechanisms including: providing regulatory elements, such as promoters and enhancers, to nearby genes, (ii) inserting themselves into genes, then targeting the epigenetic regulatory system, (iii) producing small interfering RNA. Moreover, the mobilome interacts with the rest of the genome much like the parasite-host model. Research on transposons will be helpful for the understanding genome plasticity its effect on gene expression and genome evolution.

Acknowledgements

I would like to give my sincere thanks to young researchers (Msc and PhD students) whose together with me during transposon studies. I also would like to thank Istanbul University Molecular Biology Genetics Department, Istanbul University Research Fund supporting us during our transposon research.

References


Marakli, S., Yilmaz, S., Gozukirmizi, N. 2012. BARE1 and BAGY2 retrotransposon movements and expression analyses in developing barley seedlings, Biotechnology & Biotechnological Equipment, 26: 3451-3456.


The Arabidopsis Genome Initiative 2000
The International Brachypodium Initiative 2010


Yilmaz, S., Marakli, S., Gozukirmizi, N. 2014. BAGY2 retrotransposon analyses in barley calli cultures and regenerated plantlets, Biochemical Genetics, 52: 233-244.


